CDF note 10562

Measurements of Top Quark Properties in the τ + jets decay channel at CDF

The CDF Collaboration
URL http://www-cdf.fnal.gov
(Dated: July 11, 2011)

We present a measurement of the $t\bar{t}$ cross section and the first measurement of the top quark mass in the hadronic τ + jets channel using the CDF detector at the Fermilab Tevatron. The $t\bar{t}$ cross section is derived from a Poisson likelihood function based on the number of observed and predicted events. The top quark mass is extracted using an unbinned maximum likelihood method with the probability density function evaluated for each event using leading-order $t\bar{t}$ and W + jets matrix elements and a set of parameterized jet-to-parton mapping functions. Using a total integrated luminosity of 2.2 fb^{-1}, we measure the $p\bar{p} \to t\bar{t}$ cross section to be 8.8 ± 4.3 pb^{-1}, and the top quark mass to be 172.7 ± 10.0 GeV.

I. INTRODUCTION

This note describes a measurement of the $t\bar{t}$ production cross section in $\bar{p}p$ collisions at $\sqrt{s} = 1.96$ TeV with the CDF detector at the Fermilab Tevatron, as well as the first direct measurement of the top quark mass, in τ + jets events. These measurements provide important tests of lepton universality and probe the top quark properties in a channel which may be more sensitive to new physics. Additionally, they are good examples of physics measurements performed with τ leptons in high jet multiplicity environments.

II. DATA SAMPLE & EVENT SELECTION

This analysis uses a dataset corresponding to a total integrated luminosity of 2.2 fb^{-1}collected with the CDFII detector between February 2002 and August 2007. The data was collected using a multi-jet trigger which requires at least four jets in the event each with calorimeter cluster E_T >15 GeV and a total sum E_T of all reconstructed jets >175 GeV. To these events, we apply selection cuts which require 4 jets with E_T >20 GeV, missing E_T >20 GeV, and a τ lepton with E_T >25 GeV.

Hadronically decaying τ's appear as narrow jets with an odd number of charged tracks and low π^0 multiplicity. Charged tracks are identified and measured in the central tracker, and π^0's are reconstructed from the electromagnetic calorimeter clusters with positioning information coming from hits in the CES detector. A τ candidate is seeded by a charged track with $P_T \geq 6$ GeV pointing to a calorimeter cluster. Around this seed track we define two cones. To improve jet rejection, the first cone size is variable depending on the energy of the calorimeter cluster. This signal cone extends to the minimum of 10° and (5 GeV)/E_{cal} radians where E_{cal} is the energy of the calorimeter cluster of the τ candidate. The second cone is used to define an isolation region around the τ and extends out to 30°. We reconstruct the four-momentum of the τ from all charged tracks and π^0's in the signal cone. We select τ's with $E_T \geq 25$ GeV and calorimeter cluster $E_T \geq 20$ GeV. As τ's generally have a low track multiplicity and must have an odd number of charged tracks to conserve charge, we require 1 or 3 charged tracks in the signal cone. To ensure we are selecting well isolated τ's, we require no tracks in the isolation annulus and less than 10% of the energy of the τ must present in the isolation annulus section of the calorimeter cluster. We veto electrons faking τ's requiring $E_{cal}/P^{trk} \times (0.95-\text{EMfrac}) \geq 0.1$, where E_{cal} is the total energy in the calorimeter cluster, P^{trk} is the scalar sum of the momentum of all tracks in the signal cone of the τ, and EMfrac is the fraction of calorimeter energy measured in the electromagnetic calorimeter. Finally, as the τ mass is 1.8 GeV and we cannot fully reconstruct the τ due to the presence of a neutrino, we can further reject jets faking τ's by requiring that the visible mass of the τ be less than 1.8 GeV.

Additionally, one of the 4 jets must be identified as coming from a b quark (b tagging) [1]. As our signal process only gives a single τ lepton, we also veto any event which contains an identified electron or muon.

A. QCD Multijets Removal

The dominant background for this analysis is high jet multiplicity QCD events when one of the jets fakes the signature of a τ lepton. To further reduce the QCD multijets background, we developed a neural network (NN) to distinguish between true $t\bar{t} \to \tau$ + jets events and QCD multijets events. First, we create a sample of QCD multijets events by applying the selection cuts to the data. However, we reverse one of the τ isolation cuts so that we identity jets faking τ's rather than actual τ leptons. The NN is trained to distinguish between these selected QCD multijets events and $t\bar{t}$ events, which are generated with the PYTHIA [2] Monte Carlo (MC) generator followed by the CDF detector simulation based on the Geant4 package [3]. We use 8 variables to train the NN: missing E_T, lead jet E_T, sum E_T of the jets and τ lepton, sum E_T of the two lowest E_T jets and the τ lepton, sum E_T of the two highest E_T jets, transverse momentum of the W which decays to a τ lepton, average η-moment of all jets not identified as coming from a b quark, and the lowest ratio of dijet mass to trijet mass for any possible triplet of jets.

After training the NN, we find it provides good separation between QCD multijets and $t\bar{t}$ events as can be seen in figure 1, and we chose to remove all events which return a NN output below 0.85.

III. BACKGROUND ESTIMATION

Due to the difficulty in MC modeling of QCD multijets events, b quark tagging algorithms, and the production of heavy flavor quarks in association with W bosons, we use a data-driven approach to estimate the background contribution similar to that described in [4]. First, we calculate the contributions from electroweak background

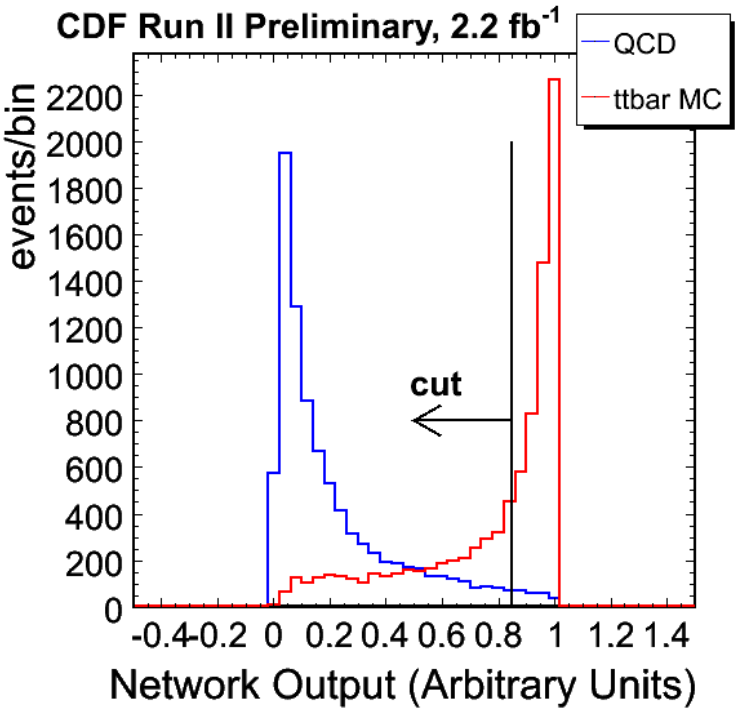

FIG. 1: Output of the neural network for signal ($t\bar{t}$) and background (QCD multijets) events after training. We select events with a NN output > 0.85.

processes and have a minimal contribution to the final total (diboson, single top quark production, and Z + jets events), as well as the signal contribution, by using the theoretical cross section for each process along with the acceptance from MC simulation and the total integrated luminosity. With these contributions known, we only need to evaluate contributions for QCD multijets and W + jets events.

To do this, we use the shape of the NN output distribution for each contribution, as well as the data, before any NN cut or b tagging requirement is applied. We then fit these distributions using a binned likelihood allowing the contribution from W + jets and QCD multijets events to float while constraining the contributions from other processes to their calculated values. From this fit, we evaluate the percentage of the data events above the 0.85 NN cut which are coming from QCD multijets. As the peak of the QCD multijets distribution is well below 0.85, the QCD multijets fit is generally derived from data outside of the signal region, and the contribution in the signal region is extrapolated from the fit.

Finally, any remaining events are assumed to come from W + jets processes. We repeat this procedure for events passing the requirement of at least one jet being b tagged. The only difference is that now the W + jets events are divided up into contributions from W + light flavor and W + heavy flavor ($W + b\bar{b}$, $W + c\bar{c}$, and $W + c$). The contribution from each process is determined by applying heavy flavor fractions and tagging efficiencies to the number of W + jets events before the b tagging requirement. The heavy flavor fraction is calculated using the ALPGEN MC generator [5] of all possible processes which may produce a W boson. Additionally, there is an MC "K-factor" applied to the contribution from heavy flavor to correct for the differences in the heavy flavor fraction observed in the data and the MC. This correction factor is derived from a NN trained to be sensitive to jets coming from heavy and light flavor quarks [4].

The contribution for each process assuming a top pair cross section of 7.4 pb and a top quark mass of 172.5 GeV is shown in Table I.

Source	Number of Events
WW	0.11 ± 0.01
WZ	0.04 ± 0.00
ZZ	0.04 ± 0.00
Stop S-chan	0.06 ± 0.01
Stop T-chan	0.10 ± 0.01
Zbb	0.29 ± 0.04
Wbb	0.57 ± 0.47
Wcc	0.34 ± 0.28
Wc	0.15 ± 0.13
W+lf	0.46 ± 0.60
QCD multijets	18.24 ± 4.10
Total Bkgd	20.40 ± 4.18
Top	18.17 ± 2.79
Total Predicted	38.57 ± 5.05
Observed	41

TABLE I: Predicted number of τ events after a neural network cut of 0.85 assuming a top cross section of 7.4 pb.

IV. TOP PAIR PRODUCTION CROSS SECTION RESULT

As the background estimate as described in section III is dependent on the top pair production cross section, we cannot simply measure the cross section as:

$$\sigma = \frac{N_{data} - N_{bkgd}}{A \cdot \epsilon \cdot \mathcal{L}}, \tag{1}$$

where N_{data} and N_{bkgd} are the number of events observed in the data and the number of estimated background events, respectively. The geometric and kinematic acceptance of the $t\bar{t}$ events is A, ϵ is the product of all the event selection data/MC scale factors (trigger, lepton identification (ID), and b-tagging), and $\mathcal{L}$ is the total luminosity. Instead, we construct a Poisson likelihood function based on the number of events in the data and the background prediction.

$$-2 \cdot lnL = -2 \cdot \left(N_{data} \cdot ln\left(\sigma_{t\bar{t}} \cdot D + N_b(\sigma_{t\bar{t}})\right) - ln\left(N_{data}!\right) - \left(\sigma_{t\bar{t}} \cdot D + N_b(\sigma_{t\bar{t}})\right)\right), \tag{2}$$

where D is the denominator of equation 1 and $N_b(\sigma_{t\bar{t}})$ is the number of events from the background prediction for a given top pair production cross section.

We calculate this likelihood for several different input top pair production cross sections ranging from 5 to 15 pb. We then fit a second order polynomial around the minimum of this function. The minimum value is taken to be the measured cross section, and the uncertainty is measured as the range of cross sections which return a likelihood result within 0.5 units of the minimum likelihood value. From the fit in figure 2, we measure the cross section to be:

$$\sigma_{t\bar{t}} = 8.8 \pm 3.3 \ (stat.) \ pb. \tag{3}$$

A. Top Pair Production Cross Section Systematic Uncertainties

We evaluate several sources of systematic uncertainties for the cross section measurement. The uncertainties considered account for systematics effects on the acceptance, selection efficiencies, background estimate, and luminosity.

For acceptance effects, we consider uncertainties on the jet energy scale (JES), initial and final state radiation, color reconnection, parton showering, and PDF uncertainties. The jet energy measured by the calorimeter is subject to several correction functions each of which give a systematic uncertainty [6]. To measure this uncertainty on the cross section, we shift the JES accordingly in the MC and re-measure the cross section. As changes in the amount of initial and final state radiation would change our acceptance, we their effect using PYTHIA MC models with increased and decreased radiation [7]. Similarly, we consider acceptance shifts from using models with and without color reconnection effects to measure this uncertainty [8]. We account for differences in parton showering models from different MC generators by performing the analysis with $t\bar{t}$ MC generated with PYTHIA and HERWIG [9]. We take the systematic uncertainty to be the difference between the two results. Finally, we consider changes in acceptance by varying the eigenvectors of CTEQ6M [10] PDF's.

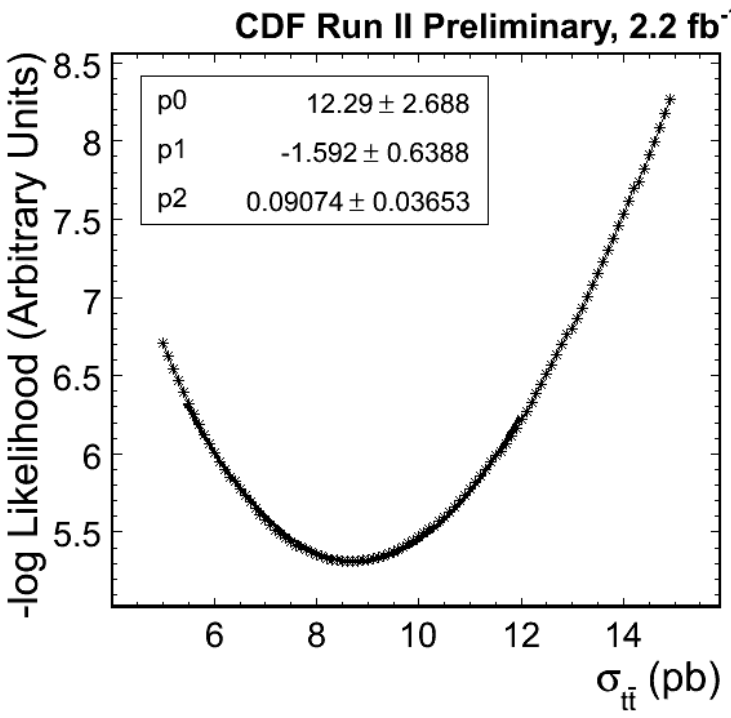

FIG. 2: Likelihood function for the top pair production cross section measurement. The likelihood is fit to a second-order polynomial (red curve) an the minimum is taken as the cross section measurement. The statistical uncertainty is determined by taking the values given by a 0.5 unit increase in the $-lnL$ from the minimum point.

We consider systematic uncertainties on the efficiency measurements on the b tagging, mistag matrix, lepton ID, and trigger. Each of these uncertainties are measured by re-measuring the cross section with the appropriate efficiency or scale factor adjusted by its systematic uncertainty. As is discussed in section III, due to inefficiencies in modeling b tagging in the MC, we use a tagging scale factor on MC jets matched to heavy flavor to account for the b tagging requirement. Similarly, due to the poor modeling of mistags in MC, we use a data based parameterization to model the mistagging of MC jets from light flavor. For the tagging and mistag matrix uncertainties, we shift the tag rate scale for jets matched to heavy flavor in the MC and mistag rate scale factor for jets matched to light flavor in the MC within its uncertainty. For the lepton ID uncertainty, we shift the lepton ID data/MC scale factor within its uncertainty and re-perform the measurement, and to measure the trigger efficiency systematic uncertainty, we shift the efficiency associated with the trigger simulation within its uncertainty.

The background systematic uncertainties are the QCD multijets contribution and the W + heavy flavor K-factor uncertainties. As QCD multijets events make up nearly 50% of our accepted data events, the contribution from QCD multijets events is our largest systematic uncertainty. To measure this uncertainty, we select multijets events from the data without the missing E_T and NN cuts. We then compare the NN output distribution of these events to that of data events with the same cuts removed both before and after the b tagging requirement. As the data events without a missing E_T requirement are QCD multijets dominated below a NN value of 0.7, we can fit the comparison between these distributions with a function to build a reweighting scheme for the QCD multijet distribution. By shifting this fit function within its uncertainty, we define a 1 σ uncertainty on the shape of the QCD multijets distribution. We then reweight the QCD multijets events and re-measure the cross section, including the multijets fitting, to measure the systematic uncertainty from the multijets contribution. For the W + heavy flavor data/MC K-factor uncertainty, we shift the K-factor from section III within its errors and take the difference in the cross section measurement as the uncertainty.

Finally, we consider the uncertainty on the luminosity measurement. The CDF detector uses Cherenkov luminosity detectors to measure the total integrated luminosity. This measurement is limited both by detector accuracy and the uncertainty on the theoretical cross section for inelastic $p\bar{p}$ collisions. As a result, the CDF luminosity has a 5.8% uncertainty associated with it [11].

For a complete list of all systematic uncertainties considered for the $t\bar{t}$ cross section analysis, see Table II.

Systematic	$\delta\sigma$ (pb)	$\delta\sigma/\sigma$ (%)
Jet Energy Scale	0.6	6.9
ISR & FSR	0.5	5.7
Color Reconnection	0.4	4.6
Tagging	0.4	4.6
Mistag Matrix	0.1	1.1
QCD Multijets Fraction	1.7	19.5
K-Factor	0.1	1.1
Parton Showering	1.7	19.5
Lepton ID	0.2	2.3
Trigger Efficiency	0.1	1.1
PDF	0.5	5.7
Luminosity	0.6	6.9
Total	2.7	31.0

TABLE II: Table of systematic uncertainties for the cross section measurement.

Having considered the various sources of systematic uncertainties, with 2.2 fb^{-1}of data we measure the top pair production cross section to be:

$$\sigma_{t\bar{t}} = 8.8 \pm 3.3\ (stat.)\ \pm 2.7\ (syst.)\ pb. \tag{4}$$

V. TOP QUARK MASS MEASUREMENT RESULT

The top quark mass measurement is derived from a likelihood function based on signal and background probabilities for each event. The method uses a similar approach as a previous measurement in the electron and muon + jets decay channels [12]. The signal probability is based on a $t\bar{t}$ leading order matrix element calculation [13] and is calculated over 31 input mass values from 145 to 205 GeVfor each event. The background probability for each event is calculated with a W + jets matrix element from the VECBOS [14] generator. Since there is no top quark mass dependence in the background probability, it is calculated only once for each event.

To improve the statistical uncertainty on the top mass measurement, we add a Gaussian constraint on the background fraction $(1 - C_s)$ to the likelihood function. The background fraction is constrained to be 0.498 ± 0.106 from Table I. The likelihood function is calculated as:

$$\mathcal{L} = \prod_{i=1}^{N} P \cdot \exp\left(\frac{-1}{2}\left(\frac{(1-c_s) - 0.498}{0.106}\right)^2\right), \tag{5}$$

where P is:

$$P = c_s P_s\left(\vec{x}; m_{top}\right) + A_{bkgd}\left(1 - c_s\right) P_{bkgd}\left(\vec{x}\right). \tag{6}$$

The signal and background probability are both calculated by integrating over the differential cross section for the appropriate process:

$$P_X = \frac{1}{\sigma}\int d\sigma(\vec{Y}) f(\tilde{q_1}) f(\tilde{q_2}) W(\vec{x},\vec{y}) d\tilde{q_1} d\tilde{q_2} \tag{7}$$

where $d\sigma$ is the differential cross section, f is the probability distribution function (pdf) for a quark with momentum $\tilde{q}$, $\vec{x}$ refers to detector reconstructed quantities, $\vec{y}$ refers to parton level quantities, and $W(\vec{(x)},\vec{(y)})$ is the transfer function used to map $\vec{x}$ to $\vec{y}$. After calculating the probabilities for each event, we evaluate a likelihood function for each of the 31 input top quark masses and fit the result with a second order polynomial to derive the central value and statistical uncertainty. For more information on the transfer functions and cross checks of the analysis method, please see the previous measurement in [12].

Before we can measure the top quark mass in the data, we must first perform a serious of validation checks and calibrations. We do this by checking the output mass versus the input mass for 31 different top quark mass points ranging from 155 GeVto 195 GeV. Additionally, we check the pull distributions for each mass point to ensure that

the central value and uncertainty are both being correctly calculated. We find that the following calibration functions are needed:

$$M_{top}^{final} = \frac{M_{top}^{measured} - 30.94}{0.806} \tag{8}$$

$$\delta M_{top}^{final} = \delta M_{top}^{measured} \times 1.76. \tag{9}$$

The large correction to the uncertainty on M_{top} is largely a result of smearing the mass distribution with the correction function 8. Because we fit the likelihood function and then apply the correction to the result, the uncertainty is unchanged by the original calibration function. This effect is then compensated for by equation 9. After applying these calibration functions, we find no bias in the residual mass or pull width (see figure 3).

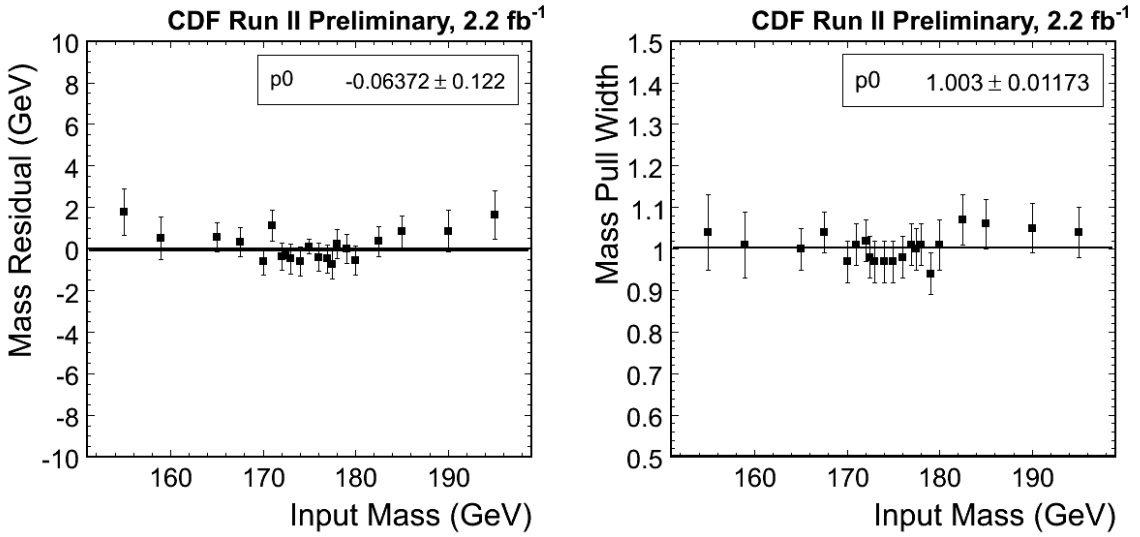

FIG. 3: Residual top quark mass (left) and top quark mass pull width (right) versus input top mass.

With the method calibrated, we measure the top quark mass in the data. The likelihood function and fit for the data can be seen in figure 4. We measure the top quark mass to be $172.7 \pm 9.3(stat.only)$ GeV

A. Top Quark Mass Measurement Systematic Uncertainties

We consider 12 different sources of systematic uncertainty for the top quark mass measurement. The largest uncertainty is from the JES. As mentioned in section IV A, a series of calibration functions are applied to calorimeter jets. Each calibration gives a corresponding uncertainty. We shift the jet energies up and down by each calibration's uncertainty and sum in quadrature the systematic uncertainty measured from each calibration function. As the top quark mass is very sensitive to the energy of its daughter particles, the JES uncertainty is the dominant uncertainty for this measurement.

We also consider systematic uncertainties from the differences in parton showering model from different MC generators, ISR and FSR, and color reconnection by performing the measurement with MC models which account for each effect. The background fraction uncertainty is measured by re-performing the measurement with MC events where the fraction of the background contribution from the QCD multijets background and each of the W + jets backgrounds is shifted within its uncertainty from Table I. The uncertainty from each background is then added in quadrature to measure the background fraction uncertainty.

To measure the uncertainties from PDFs, we account measure the shift from the different CTEQ6 eigenvector PDFs. To account for the uncertainty on the fraction of gluon-gluon fusion produced $t\bar{t}$ events, we reweight the MC events so that the percentage of $t\bar{t}$ events which result from gluon-gluon fusion is shifted from 5% to 20%.

The b-jet uncertainty accounts for different fragmentation models and semileptonic branching ratios for jets from b quarks [15]. These uncertainties are added in quadrature to an uncertainty measured from shifting the energy scale

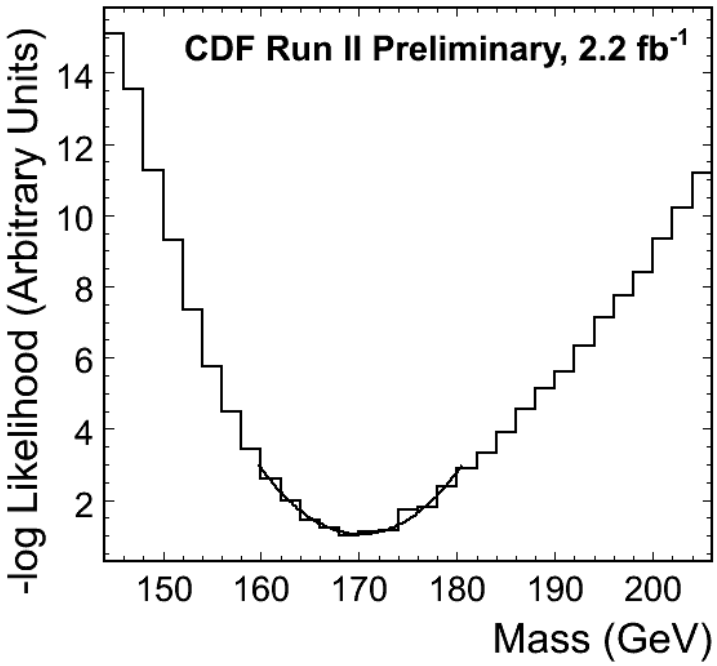

FIG. 4: Negative log likelihood function for the top quark mass measurement. The black curve shows the fit to a second order polynomial.

Source	Result (GeV)
JES	3.37
MC Generator	0.50
ISR/FSR	0.34
Color Reconnection	0.50
Background Fraction	0.47
MC Statistics	0.14
PDF	0.12
gg fusion	0.17
B-jet	0.39
Lepton p_T	0.19
Pileup	0.95
Calibration	0.17
Total	3.7

TABLE III: Total Systematic Uncertainties for tau channel.

of jets from b quarks to get the total b-jet uncertainty. We also account for shifts from the lepton energy scale by considering changes in the measurement from MC samples with shifted τ energy.

The pileup systematic uncertainty accounts for a known mismodeling in the luminosity profile of the MC. To account for this, we measure the shift in the measurement from MC events which are reweighted to give the luminosity profile which is seen in the data.

Finally, we consider systematic uncertainties from our calibrations. We shift the calibration function in equation 8 within its uncertainty to measure the systematic uncertainty from the calibration function. Even after the calibration function is applied, we find a 0.14 GeVuncertainty on the fit of the mass residual across all 31 mass points. Due to this, we take a 0.14 GeVsystematic uncertainty for MC statistics.

The full table of systematic uncertainties for the top quark mass measurement can be found in Figure III.

Having considered the various sources of systematic uncertainties, we find the top quark mass measured directly in

events from the tau + jets decay channel with 2.2 fb^{-1}of data to be:

$$172.7 \pm 9.3\ (stat.)\ \pm 3.7\ (syst.)\ \text{GeV}. \tag{10}$$

VI. CONCLUSIONS

We use the τ + jets decay channel to identify $t\bar{t}$ events as well as measure the top quark properties with 2.2 fb^{-1} of data. We find the $t\bar{t}$ pair production cross section to be $8.8 \pm 3.3\ (stat.) \pm 2.7\ (syst. + lumi.)$ pb or 8.8 ± 4.3 pb. We also measure the top quark mass in this decay channel for the first time ever. With τ + jets events, we find the top quark mass to be $172.7 \pm 9.3\ (stat.) \pm 3.7\ (syst.)$ GeV or 172.7 ± 10.0 GeV. We find the measurements to be consistent with the current world average top pair production cross section of 7.5 ± 0.5 pb and the current world average top quark mass of 172.70 ± 1.09 GeV. As the values we measure with tau leptons agree with current measurements, this analysis is an important test of lepton universality.

Acknowledgments

We thank the Fermilab staff and the technical staffs of the participating institutions for their vital contributions. This work was supported by the U.S. Department of Energy and National Science Foundation; the Italian Istituto Nazionale di Fisica Nucleare; the Ministry of Education, Culture, Sports, Science and Technology of Japan; the Natural Sciences and Engineering Research Council of Canada; the National Science Council of the Republic of China; the Swiss National Science Foundation; the A.P. Sloan Foundation; the Bundesministerium fuer Bildung und Forschung, Germany; the Korean Science and Engineering Foundation and the Korean Research Foundation; the Particle Physics and Astronomy Research Council and the Royal Society, UK; the Russian Foundation for Basic Research; the Comision Interministerial de Ciencia y Tecnologia, Spain; and in part by the European Community's Human Potential Programme under contract HPRN-CT-20002, Probe for New Physics.

[1] T. Affolder *et al.*, Phys. Rev. D **64**, 032002 (2001).
[2] T. Sjostrand *et al.*, High-Energy-Physics Event Generation with PYTHIA 6.1, Comput. Phys. Commun. **135**, 238 (2001).
[3] Geant 4 Collaboration, S. Agostinelli *et al.*, Geant4 A Simulation Toolkit, Nucl. Inst. Meth., A **506**, 250, (2003).
[4] T. Aaltonen *et al.*, Phys. Rev. Lett. **105**, 012001 (2010).
[5] M. L. Mangano *et al.*, J. High Energy Phys. 07, **001**, (2003).
[6] A. Bhatti *et al.*, Nucl. Instrum. Meth., A **566**, 375 (2006).
[7] A. Abulencia *et al.*, Phys. Rev., D **73**, 032003 (2006).
[8] D. Wicke and P. Z. Skands, Eur. Phys., J **C52**, 133 (2007).
[9] G. Corcella *et al.*, J. High Energy Phys., **01**, 010 (2001).
[10] J. Pumplin *et al.*, J. High Energy Phys. 07, **012**, (2002).
[11] V. Papadimitriou eprint arXiv:1106.5182
[12] A. Abulencia *et al.*, Phys. Rev. Lett. **99**, 182002 (2007)
[13] G. Mahlon and S. J. Parke, Phys. Rev. D **53**, 4886 (1996)
[14] F.A. Berends, H. Kuijf, B. Tausk and W.T. Giele, Nucl. Phys. B **357**:32-64 (1991)
[15] T. Aaltonen et al., Phys. Rev. D **79**, 092005 (2009).